

10/670,998 AU-892

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 January 2006 (19.01.2006)

PCT

(10) International Publication Number  
WO 2006/005187 A1

(51) International Patent Classification : G06T 7/00, 7/60, G06F 17/30

(21) International Application Number: PCT/CA2005/001093

(22) International Filing Date: 8 July 2005 (08.07.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 60/586,176 9 July 2004 (09.07.2004) US

(71) Applicants and

(72) Inventors: AARABI, Parham [CA/CA]; 909 Bay Street, Apt. 1505, Toronto, Ontario M5S 3G2 (CA). MAVANDADI, Sam [CA/CA]; 214 Bogert Avenue, Toronto, Ontario M2N 1L3 (CA).

(74) Agent: GIERCZAK, Eugene J.A.; Miller Thomson LLP, 20 Queen Street West, Suite 2500, Toronto, Ontario M5H 3S1 (CA).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AB, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

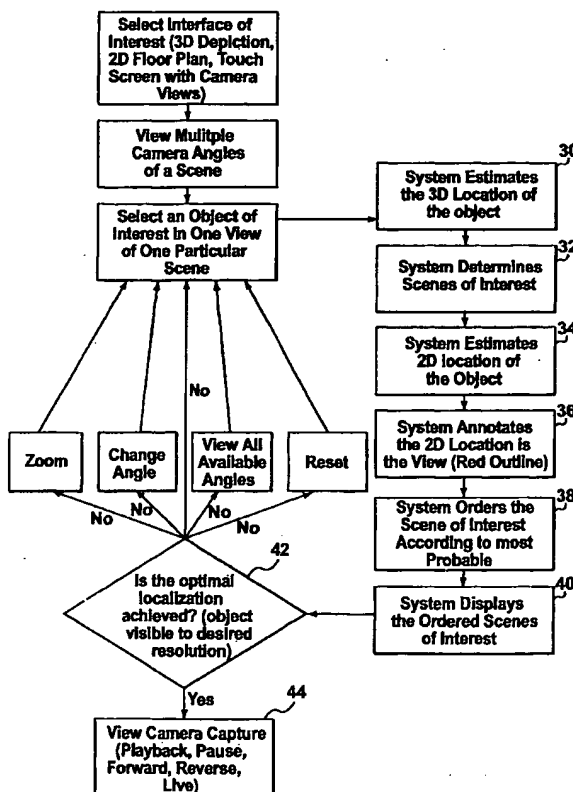
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

[Continued on next page]

(54) Title: INTERACTIVE THREE-DIMENSIONAL SCENE-SEARCHING, IMAGE RETRIEVAL AND OBJECT LOCALIZATION



(57) Abstract: An interactive image searching and sorting method, system and computer program is provided that makes use of the three-dimensional relations among a plurality of images. Calibrated cameras are used to capture images from a single scene. The user interacts with the system by means of selecting a point of interest on an image. The system generates a probabilistic model of the user's selection and determines the location of the object of interest with the use of spatial likelihood and reliability functions that model the likelihood of the object location and error in sensors. The system iteratively, based on user input, processes the images and based on the probabilistic model ranks the images to provide the 'best' view of the region of interest.

WO 2006/005187 A1



- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## INTERACTIVE THREE-DIMENSIONAL SCENE-SEARCHING, IMAGE RETRIEVAL AND OBJECT LOCALIZATION

### Background of the Invention

5           With the explosion of images on the World Wide Web and creation of online databases for visual media, the ability to search through large numbers of images has become evermore important. As a result, a great deal of research has been done in the field of image searching. Most existing algorithms for image searching and retrieval use some form of content-based algorithm in order to accomplish the goal, as particularized in  
10   References [1]–[4] below. Many techniques use a color-histogram to analyze and compare the color contents of different images and sort them accordingly, as set out in Reference [5] below. Other methods apply computer vision and image processing algorithms such as segmentation as per Reference [6] below on the different images, and attempt to extract the shape and contents of the components of the images. In the case of searching for images on  
15   the web, since there is not a great deal of time to complete the search, most searching algorithms, search the content of the pages that are linked to the image for keywords. For example, the authors in Reference [7] below develop a method for weighing the available side-information for this purpose. There are still other methods, for example, in Reference [8] below, the authors suggest an interactive system with a neural network to accomplish the  
20   task. It should also be mentioned that more recent developments such as those outlined in References [2], [7]–[9] below all use some sort of interactive or relevance-based method to increase the accuracy and robustness of the system.

          The algorithms disclosed in the prior art references listed above generally do not assume three dimensional dependence among the images, such that when comparing images  
25   there is no difference in the processing of the data whether the images are taken from the same scene, or from different scenes. If it is known that the images are all from the same general location, then other features can be used to compare the images. More specifically, if the images are taken using cameras with known calibration parameters, significant geometric and spatial data can be obtained that can be used for comparisons and sorting.

30           What is needed is a method, system and computer program that enables images to taken from the same scene to be searched and sorted in an efficient manner.

### Summary of the Invention

One aspect of the present invention consists of a method, system and computer program for ranking images taken from a single scene based on the visual coverage that the images provide for a particular object or region in space. In a particular aspect of the invention, first it is determined whether the object or region is present or not in the images. Therefore, the ranking in a first stage distinguishes among images that include and those that do not include the desired object or region. Second, from the remaining images those images are separated that have a 'better' view of the particular object or region. In a particular aspect of the present invention, means is provided for selection of images that have this "better" view.

In another aspect of the present invention, a method of searching scenes in images, retrieving images, and/or localizing objects in images, characterized by: obtaining a plurality of images from one or more cameras, the plurality of images including at least two images including a view of a single area or object shown in a scene; selecting a particular area of interest or object of interest in the scene; and iteratively establishing a sub-set of the plurality of images that are probably of interest for viewing the particular area of interest or object of interest by: (i) determining a probability distribution of the plurality of images based on location data and data regarding the geometry of an environment of the scene established for the area of interest or object of interest; (ii) refining the probability distribution by obtaining user input regarding: (a) one or more of the sub-set of plurality images that the user considers to be most relevant from the current sub-set of the plurality of images; and (b) selection of the particular area of interest or object of interest in the one or more most relevant images; and (iii) updating the sub-set of the plurality of images based on the user input of (ii).

### Brief Description of the Drawings

A detailed description of several embodiments of the present invention is provided herein below by way of example only and with reference to the following drawings, in which:

Figure 1 is a system resource diagram illustrating the principal resources of the system of the present invention, in accordance with one particular embodiment thereof.

Figure 2 is a program resource diagram illustrating the principal resources of the computer program of the present invention, in accordance with one particular embodiment thereof.

Figure 3 is a process flowchart illustrating the steps of the method of the present invention,  
5 in accordance with one particular embodiment thereof.

Figure 4a is an illustration of a representative interface for the computer program of the present invention, in a particular aspect thereof.

10 Figure 4b is an illustration of another representative interface for the computer program of the present invention, in a particular aspect thereof.

Figure 5a is a diagram that illustrates the method for identifying points within the Field of View or FOV.  
15

Figure 5b is a diagram that illustrates the method for determining whether a point is within the FOV.

Figure 6 illustrates a multi-camera system with overlapping FOV's.  
20

Figure 7 illustrates a series of images taken from a scene that include a plurality of images, the images including a plurality of images that show a single image in a plurality of views thereof. The images are in a random order.

25 Figure 8 illustrates the selection of the particular object and the sorting of the images in order of relevance to providing views of the particular object, in accordance with the present invention.

Figure 9 illustrates fifteen highest-ranked images after selecting from the recycle bin twice,  
30 in accordance with the present invention.

Figure 10 shows the highest ranked images after selecting from the recycle bin three times, in accordance with the present invention.

In the drawings, preferred embodiments of the invention are illustrated by way of example.

- 5 It is to be expressly understood that the description and drawings are only for the purpose of illustration and as an aid to understanding, and are not intended as a definition of the limits of the invention.

#### Detailed Description of the Invention

- 10 Fig. 1 illustrates the system of the present invention, in one particular embodiment thereof. Typically the system includes, or is linked to, a camera network or camera array (10). As explained below, the present invention enables scene searching, image retrieval and object localization in relation to a plurality of images, in which the plurality of images include at least two images that include views of a particular object or scene taken by a
- 15 camera. While the present invention can be practiced in relation to such images taken by a single camera, typically the invention is practiced in relation to a plurality of cameras linked in the camera network or camera array (10) depicted in Fig. 1.

- Depending on the type of cameras used in the camera array (10), if the camera array generates digital images, then the camera array (10) is typically linked to an IP network (12),
- 20 and the digital images are stored to the digital archive (12). If the camera array (10) generates analog images, then the camera array (10) is linked to an analog network (14) and the image recording is converted (16) by operation of a suitable analog to digital converter, and then the resulting digital images are stored to the digital archive (12).

- The computer system of the present invention is generally illustrated as a
- 25 Computation Means (20) in Fig. 1, which in a typical implementation of the present invention consists of the computer program (22) (best understood by reference to Fig. 2) of the invention loaded on a computerized device (not shown) linked to the camera array (10).

- In a particular embodiment of the present invention, the principal resources of the computer program of the present invention are illustrated in Fig. 2. For the sake of clarity,
- 30 Fig. 2 illustrates the present invention in representative blocks, namely the interface block

(24), the storage block (26) and the computation block (28) for the sake of understanding the principal functions of the computer program. The organization of the computer program (22) into blocks (24), (26), (28) should not be understood as referring to a particular computer program structure and therefore limiting in any way the present invention to a particular computer program, or particular structure thereof. The functions of the computer program of the present invention can be provided in more or less, or different blocks than as illustrated in Fig. 2.

The functions of the computer program (22) are explained in greater detail below, including by reference to Fig. 2.

#### 10 *Outline of Method Steps*

The method of the present invention is best understood by reference includes following steps (and Fig. 3 illustrates a particular embodiment of this method, as explained in greater detail below):

1) *Obtaining Images*: a plurality of images are obtained, the plurality of images including at least two images including view of a single area of a scene or object shown in the scene. Generally, the present invention assumes a relatively large number of images, which require searching and sorting to derive one or more images comprising a subset of the universe of images obtained, as particularized below. These images are obtained from the digital archive (12), as particularized above. For the sake of clarity, the images are obtained from the camera array (10), typically consisting of a relatively large array of video-cameras or they can be still images of a particular environment taken with a single or multiple cameras. The images are generally assumed to be available to a user in random order.

2) *Calibration*: In order to process the available images using the three dimensional information from the environment, they are preferably captured using calibrated cameras. The cameras of the camera array (10) are generally calibrated prior to the capture of the images referred to in 1) above, however, as particularized below in 5) the present invention involves, where necessary, further calibration of the cameras of the camera array (10) in response to the search/retrieval/localization functions described below. Therefore in a particular aspect of the present invention, the camera array (10) is responsive to a series of calibration commands from the computer program (22) in conjunction with the

search/retrieval/localization functions of the present invention. As an alternative to such calibration, images can be captured and then cameras calibrated using landmarks in a manner that is known.

3) *Object Selection*: It is assumed that the user is interested in a particular object or region in space depicted in at least two of the images obtained. The computer program (22) is operable to take as input the selection of a point on an image within the presented set of images. The selection occurs by operation interface illustrated in Fig. 4a by which a user selects an image point, typically using a cursor.

4) *Probabilistic Localization*: The image point that the user selects corresponds to a line in three dimensions. Moreover, it is assumed that the user is not interested in a single point in space, but rather a region. Therefore it is assumed that the user is interested in a region of the image, with a certain probability distribution as further explained below. Using this distribution and the available camera views, the images can be sorted and the location of the region of interest can be narrowed down.

5) *Refinement*: The last three steps are repeated until a reasonable degree of localization is achieved as particularized below.

The steps 1) to 5) above are explained in greater details below.

#### ***Camera Geometry and Calibration***

Camera calibration is a subject that has been extensively explored in the literature, for example Reference [10]–[22] below. The below provides a camera calibration in the context of the present invention, and in particular calibration as described in step 2) above.

#### ***Homogeneous representation***

Any line on a plane can be represented by an algebraic equation of the form  $\alpha x + \beta y + \gamma = 0$

The different choices of  $\alpha$ ,  $\beta$  and  $\gamma$  determine the direction of the different lines. Thus a line on a plane can be represented by a vector of the form  $(\alpha, \beta, \gamma)^T$ . Any multiple of this vector represents the same line, and the set of all of these multiples form an



equivalence class called a *homogenous* vector. The set of equivalence classes in  $\mathbb{IR}^3$  forms the *projective space*  $\mathbb{IP}^2$ .

Equation (1) also represents all the points  $\mathbf{x} = (x, y)$  that lie on the line  $\mathbf{l} = (\alpha, \beta, \gamma)^T$ . This equation can be written as a dot-product:  $(x, y, 1) \cdot \mathbf{l} = 0$ , which is unique up to a scale factor. As a result, any point in  $\mathbb{IR}^2$  can be represented as  $\kappa(x, y, 1)$ , with  $(x, y)$  being the actual co-ordinates in  $\mathbb{IR}^2$  and  $\kappa$  a constant scalar. This is the homogeneous representation of a point on a plane, and is an element of the projective space  $\mathbb{IP}^2$ . Similarly a point in 3D-space is denoted by a 4-element vector of the form  $\mathbf{X} = \xi(X, Y, Z, 1)^T$ , where  $\xi$  is a constant scalar. Note that in the disclosure below, small bold-faced letters such as  $\mathbf{x}$  are used to denote 2D image points, and capital bold-faced letters such as  $\mathbf{X}$  used to denote 3D world points.

### **Calibration**

The cameras that are part of camera array (10) are generally assumed to be general projective cameras (for example as particularized in Reference [10] below) and are generally calibrated using linear methods in a manner that is known.

When an image is captured by a camera of the type particularized, there is a transformation that takes the three dimensional coordinates of points in the real world and maps them onto a two dimensional image plane. This transformation is described by the following equation:

$$\mathbf{x} = \mathbf{P}\mathbf{X}$$

where  $\mathbf{X}$  is the coordinates of a point in the real world, and  $\mathbf{x}$  is the corresponding coordinates on the image plane.  $\mathbf{P}$  is the *Projection Matrix* that transforms the 3D points to 2D. As mentioned before, the  $\mathbf{x}$  and  $\mathbf{X}$  are represented using homogeneous coordinates, and so  $\mathbf{P}$  is a  $3 \times 4$  matrix. Calibrating a camera as described in the present invention consists generally of calculating the elements of this matrix. The matrix can be decomposed to extract intrinsic and extrinsic parameters of the camera such as the focal length, skew, rotation, translation, etc. In the next section, the method for computing  $\mathbf{P}$  as it is shown in Reference [10] is discussed.

**Solving for  $P$** 

In order to solve for the projection matrix, we represent  $P$  in the following manner:

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix} = \begin{bmatrix} p^{1T} \\ p^{2T} \\ p^{3T} \end{bmatrix}$$

where  $P^{kT}$  is the  $k^{\text{th}}$  row of the matrix  $P$ .

- 5 Now, it is desired to solve for  $P$  such that  $\mathbf{x}_i$ , where  $i$  indicates the  $i^{\text{th}}$  known point used for calibration and  $\mathbf{x}_i = (x_i, y_i, 1)^T$  is the image coordinates of the real world point  $\mathbf{X}_i$ . Since homogeneous coordinates are being used, everything is unique up to a scale factor. Hence,  $\mathbf{x}_i$  and  $P\mathbf{X}_i$  are only in the same direction, and not necessarily equal. This leads to the following cross-product equality:

$$10 \quad \mathbf{x}_i \times P\mathbf{X}_i = 0$$

Then using (3)

$$\mathbf{x}_i \times P\mathbf{X}_i = \begin{pmatrix} y_i P^{3T} \mathbf{X}_i - w_i P^{2T} \mathbf{X}_i \\ w_i P^{3T} \mathbf{X}_i - x_i P^{2T} \mathbf{X}_i \\ x_i P^{3T} \mathbf{X}_i - y_i P^{2T} \mathbf{X}_i \end{pmatrix}$$

Now,  $(P^{kT} \mathbf{X}_i = \mathbf{X}_i^T P^k$  for  $k=1, \dots, 3$ , which results in the following:

$$\begin{bmatrix} 0^T & -w_i \mathbf{X}_i^T & y_i \mathbf{X}_i^T \\ -w_i \mathbf{X}_i^T & 0^T & -x_i \mathbf{X}_i^T \\ -y_i \mathbf{X}_i^T & x_i \mathbf{X}_i^T & 0^T \end{bmatrix} \begin{pmatrix} P^1 \\ P^2 \\ P^3 \end{pmatrix} = 0$$

- 15 Note that  $w_i$  are the scale factors for  $\mathbf{x}_i$ . Now let  $\mathbf{p}$  be the vector containing the entries of the matrix  $P$  and  $A$  be the leftmost matrix in (4c). Consequently, it is necessary to effectively solve for the equation  $A\mathbf{p}=0$ . Note that only two of the rows of the matrix  $A$  are linearly independent, thus to find a solution, at least six points are needed, since there are 11

degrees of freedom in the matrix  $P$  and each point of correspondence results in two linearly independent equations. It must however be noted that it is only in the absence of noise that the point correspondences are 'perfect' and using six points to find a solution results in the unique and correct solution. In practice, however, the measurements are not perfect and as a result generally many point correspondences are required to solve the system. These correspondences (more than six) result in an over-determined system that must be solved while minimizing some error-measure. A standard approach is to minimize  $\|Ap\|$  under the constraint that  $\|p\|=1$  (the norm of  $p$  is of no consequence since  $P$  is only defined up to a constant scale factor) [10]. This problem is the same as finding the minimum of  $\|Ap\|/\|p\|$ . The solution is the unit eigenvector of  $A^T A$  with the least eigenvalue. This is the same as the unit singular vector corresponding to the smallest singular value of  $A$ .

#### Spatial Likelihood Functions (SLFs)

One of the aspects of the present invention, as particularized below, involves localization of the particular object or scene. This localization depends, in a particular embodiment of the present invention, on localization as described under this heading.

In most practical settings, it is extremely difficult, if not impossible to localize an event or phenomenon deterministically and perfectly accurately using sensors. As a result, a probabilistic approach is preferred as compared to a deterministic one. Under these conditions, preferably the following are obtained: the probability mass function of  $X$ , the object location, or some function that is proportional to  $P(X)$ . It should be understood, however, that for the purposes of localization, only a monotonically decreasing function as opposed to an actual distribution.

$$\Gamma X = \psi(P(X = X_w | \Theta))$$

where  $\Gamma(X)$  is the SLF at spatial location  $X$ ,  $X_w$  is the true location of interest,  $\Theta$  represents all available data, and  $\psi(t)$  is a monotonically decreasing function of  $t$ .

**SLF Generation**

It is assumed in this part of the disclosure that all of the cameras forming part of camera array (10) have been perfectly calibrated. Thus their corresponding calibration matrices  $P_i$ , corresponding to the  $i^{th}$  camera are available.

5       The point  $x_{u,i}$  corresponds to a line from the centre of the camera and through the selected point on the image plane, intersecting the selected object/point in 3D space. It is assumed that the objective is to select a region in space or an object of finite size as opposed to a single point in space. There is also a certain amount of error associated with the point selected by the user. The selection is done in a probabilistic manner. By looking at the image  
 10       plane and taking the selected point as the mean of a Gaussian distribution with a user-defined variance, a region of space around the mean is selected, with the size determined by the variance. Thus every point that lies on the image plane, regardless of whether it is in the FOV of the camera or not, will have a likelihood value associated with it. This is the likelihood that the point is the point of interest. Now, since every point in space as seen by  
 15       the camera lies somewhere on the image plane, it will have an associated probability value. This way the likelihood of every line through the camera centre has been determined. Note that the variance of the Gaussian distribution determines how large of a volume is desired for the final localization, and this should be varied for different applications.

To generate the SLF, the volume of space to be considered is first determined. This  
 20       can be taken to be a cube of length  $L$ , with  $C_1$  at the centre of one face 'looking' inside the cube. Now let  $\chi$  be the set of all points in this volume of space. Then, likelihood values can be assigned to one such cube corresponding to the region of interest as observed by the first camera by looking at the projection of each point in the defined space on the image plane. Note that for successive SLFs, the same  $\chi$  that was defined with the first camera located at  
 25       the centre of one face is considered; this however does not have to be the case and as long as the volume is a fixed one, the system will work properly.

Thus the projection of the SLF on the image plane will have the following form:

$$P(\mathbf{x} = \mathbf{x}_{u,i}) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2} \left[ \left( \frac{x - x_{u,i}}{\sigma_x} \right)^2 + \left( \frac{y - y_{u,i}}{\sigma_y} \right)^2 \right] \right\}$$

Where  $\mathbf{x}_{u,i} = (x_{u,i}, y_{u,i}, 1)^T$ ,  $\mathbf{x} = P_i \mathbf{X} = (x, y, 1)^T$  and  $\mathbf{X}$  is the 3D coordinates of a point in the volume of space to be analyzed. Now  $\sigma_x$  and  $\sigma_y$  determine the size of the volume that the system is focusing on and may vary depending on the required search resolution.

5 Without loss of generality, it can be assumed that these two values are equal.

It must be noted that this SLF never assumes zero as a value. The Gaussian decays to zero at infinity. This also allows for easier computation of the true 3D SLF since in general, there will be points in  $\chi$  that are not seen by the camera under consideration. Projecting these points using  $P_i$  results in pixel coordinates outside the resolution of the camera, and  
 10 thus a low probability that the point is of interest to the user.

#### Spatial Reliability Functions (SRFs)

Another aspect of the present invention involves assessing whether a particular point lies in the FOV of a particular camera and how much reliable access the particular camera has to the particular point. This is determined using assessment of the Spatial Reliability  
 15 Functions of particular cameras forming part of the camera array (10), as explained below.

Whenever multiple sensors are used in an environment to gather information regarding an event or a phenomenon at some location, the data obtained using each sensor has a certain level of reliability associated with it. This level of reliability may be due to the proximity of the sensors to the phenomenon of interest, the intrinsic properties of the  
 20 sensors, or other factors that may be caused by the structure of the sensor network and the environmental setting. For example, the data obtained from an acoustic sensor (e.g. microphone) closer to a sound-source is more reliable (has higher signal to noise ratio) than that obtained from a sensor which is far away. Therefore, for every sensor in the system, a probability value can be assigned to every point in the space that is of interest. This is a  
 25 spatial reliability function, and for each sensor, it represents the likelihood of the reliability of any information obtained regarding a specific spatial coordinate. If it has a value of unity

at a specific location, then this means that the information obtained regarding that location is perfectly accurate and is not corrupted by any type of noise; and if it has a value of zero, then the information obtained is completely inaccurate.

### *SRF Generation*

- 5 To generate an SRF for the  $i^{\text{th}}$  camera, we first find the set of all points  $\beta$  that are both in  $\chi$  and in the FOV of the  $i^{\text{th}}$  camera  $\mathcal{F}_i$ .

$$\mathcal{F}_i = \left\{ X = (x, y, z) : X \in \text{FOV of the } i^{\text{th}} \text{ camera} \right\}$$

$$\beta_i = \left\{ X = (x, y, z) : X \in \{ \chi \cap \mathcal{F}_i \} \right\}$$

- 10 Then reliability values are assigned to each point in  $\beta$ , according to a monotonic radial decay. This is the SRF for the  $i^{\text{th}}$  camera  $\rho_i(X)$ , where the maximum is at the camera center (slightly in front of the camera in fact).

### *Identifying points within the FOV*

- There are a number of methods that could be used to determine  $\mathcal{F}$ . One method is to find the lines that originate from the centre of the camera and go through the four corners of the image plane and then use the planes spanned by adjacent pairs of these lines as boundaries of the FOV. This is shown in Figure 5(a). One first has to find the equation of the lines. Two points on each line are known, and those yield an equation. Take  $l_j, j \in \{1, 2, 3, 4\}$  to be a line going through one of the corners, and  $\mathbf{x}_{c,j}$  to be the corresponding corner point on the image plane. The equation of  $l_j$  then becomes [10]:

20 
$$X_j(\lambda) P^\dagger \mathbf{x}_{c,j} + \lambda C$$

here  $C$  is the coordinates of the camera, and  $X_j$  corresponds to any point that lies on the line  $l_j$ .  $P^\dagger$  is the pseudo-inverse of the projection matrix  $P$  and is defined as the following:

$$P^\dagger = P^T (PP^T)^{-1}$$

thus  $PP^\dagger = I$ . This leads to the following projection of points on the line  $l_j$ :

$$\begin{aligned}
 P X_j(\lambda) &= PP^T x_j^c + \lambda PC \\
 &= x_j^c + \lambda \cdot 0 \\
 &= x_j^c
 \end{aligned}$$

- since the coordinates of the camera when projected using  $P$  maps to  $(0,0,0)^T$ . Now, any point on this line maps onto the same point on the image plane. To determine whether a point is in the FOV or not, it is determined if the point lies in front or behind the camera. Any point  $X = (X,Y,Z,T)^T$  in space lies in front of the camera if it has a positive depth [10] defined by:

$$\text{depth}(X; P) = \frac{\text{sign}(\det M)w}{T\|m^3\|}$$

- where  $P = [M|p_4]$  is the projection matrix for the camera, and  $P = (X,Y,Z,T)^T$ . And  $m^3$  is the third row of  $M$  and thus a vector in the positive axial direction. To find whether a point is actually in the field of view of the camera, the planes shown in Figure 5(b) are established. Then angle between the line connecting the point in question to the camera centre and each of the planes can be determined. Knowing the planes  $P_1, P_2, P_3$  and  $P_4$  each visible point can only be in a certain range of angles from the planes, and its can be determined whether the point is in the FOV of the camera.

- Although the technique described above works well, it is not the easiest method for practical purposes. The easiest way to determine whether a point lies in the FOV would be to project it onto the image plane using  $P$  and see if it is within the image resolution of the camera. For example, for a camera with resolution  $640 \times 480$ , if the projected point is in  $[0, 639]$  in the horizontal direction and  $[0, 479]$  in the vertical direction, then the point is in the FOV. Note that the absolute values in the range depend on the pixel which is taken to be the origin. Therefore, in order to find  $\beta$ , we can take every point in  $\mathcal{F}$  and see whether it lies in the appropriate range of the image plane coordinates, and if so then the point is also in  $\beta$ . Also note that we still have to use the depth function to determine whether the point is in front or behind the camera.

***Viable SRFs for cameras***

It can be assumed that the ability of a camera in 'observing' objects decreases with distance. Therefore the SRF must decay monotonically with distance from the camera center. The rate of the decay depends on the resolution of the camera and more generally on the 'quality' of the camera and may be determined experimentally. A possible SRF definition is given below:

$$\rho_c(X) = \begin{cases} \max(0, \kappa \cdot e^{-\gamma \cdot \|X-C\|}) & \text{if } X \in B \\ 0 & \text{if } X \notin B \end{cases}$$

where  $X \in \mathcal{F}$ ,  $C$  is the coordinates of the camera, and  $\gamma$  is determined experimentally.  $\rho_c$  is the SRF.  $\kappa$  is a constant that determines how much more the closer regions are emphasized as compared to the regions farther away and  $\gamma$  is the decay rate, determined experimentally. This exponential decay allows the system to emphasize regions of space that are closer to the camera and punish those regions farther away. As a result, cameras that are closer to the region of interest will be considered more reliable.

In the experiments conducted here, an exponential SRF has been used, with  $\kappa = 30$  and  $\gamma = 30\text{cm}^{-1}$ .

**Further Description of the Method**

In accordance with a particular aspect of the present invention, the camera array (10) includes at least two cameras having overlapping FOVs, as shown in Figure 6. So, any given object, or region in space, lies within the common FOV of more than one camera in this particular arrangement. Therefore, images of the object or region of interest are available from different angles and different positions, i.e. images are taken from different distances from the region. Using this, we can ignore the irrelevant images, i.e. those obtained using cameras whose FOVs do not contain the desired spatial coordinates. Furthermore, we can rank the remaining images, based on the distance of their corresponding cameras from the region of interest, and whether they completely contain the desired spatial coordinates or not. This is accomplished by assigning an SRF to each camera and generating an SLF, as described above. The computer program of the present invention including computer



instructions that when provided to a computerized device are operable to provide this function.

A representative algorithm is described which when provided as computer instructions forming part of the computer program of the present invention, provides the function particularized. It should be understood that other algorithms providing this function are possible.

The algorithm initially generates the SRFs for all the cameras using Equation 11. After the user selects the first view and the region of space within that view, the corresponding SLF is generated and is combined with the SRFs just by pointwise multiplication and normalization. To do this, a normal distribution with mean at the selected point is assigned to pixels in the image using Equation 6 representing the probability of the pixels being the point of interest on the image. Then every point in the environment is projected onto the image plane of the active camera using Equation 2. The likelihood of any point in the environment being of interest is set to be equal to the likelihood value associated with its projection onto the 2D image plane. It must be noted that this is an interactive localization based on feedback from the user, and the SLFs do not correspond to the same location after each iteration; they instead correspond to lines and regions in space that should ideally have an intersection at the point of interest.

To navigate through the available camera viewpoints, there is a need to assign degrees of validity to each camera forming part of the camera array (10), for the particular selected scene. Immediately after the first selection, one SLF and all the SRFs for the cameras are available. According to Reference [23] cited below,  $E_r(x)[\rho_i]$  can be used instead of  $E_r(x)[\rho_i]$  as a measure of the reliability. Therefore, to rate each remaining camera in the camera array (10), one can multiply the SLF with each SRF and sum all the likelihood values. This gives a measure of the level of access that each camera has of the selected region. In other words, it shows how well each camera can observe the selected region. The user now moves to another view and selects another point. Again an SLF is generated, and this time it is multiplied with the previous SLF and normalized; this further narrows down the region of interest, and again the remaining camera views are ranked based on this combined SLF and their SRFs. The process is continued until the desired view is found. Alg. 1 shows step by step how ISL works.

- The spatial coordinates of the location of the object of interest can be estimated by taking the expected value of the SLF when the SRF is factored in. This means that to find the location of the object or region of interest in using a particular camera, the SLF is multiplied with the SRF and then the maximum coordinates are projected onto the image plane. The accuracy of this estimate increases as the number of selections by the user is increased. This is because after each selection, the SLF gets more and more concentrated in a very small volume, which is common among all the individual SLFs corresponding to each selection from a particular viewpoint.

---

**Algorithm 1**


---

- 1: Generate SRFs for all cameras  $C_i$ :

$$\rho_i(X) = \begin{cases} \max(0, \mathcal{K} \cdot e^{-\gamma \cdot \|X - C_i\|}) & \text{if } X \in \mathcal{B} \\ 0 & \text{if } X \notin \mathcal{B} \end{cases}$$

- 2: Allow user to select an initial viewpoint  $\mathcal{V}_j$ ,  $j = 0$ .  
 3: Allow user to select an initial point  $x_{u,j}$  from  $\mathcal{V}_j$ .  
 4: Based on the environment select the volume of space  $\chi$  to analyze.  
 5: Compute  $P(X_{u,j} = X)$ , the SLF for the initial point using the projection of the SLF:

$$P(x = x_{u,i}) = \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{1}{2\sigma^2} \left[ (x - x_{u,i})^2 + (y - y_{u,i})^2 \right] \right\}$$

- 6: Let  $\Gamma(X) = P(x_{u,j} = X)$  be the most recently obtained SLF.  
 7: Compute the expectation of the SRF according to  $\Gamma(X)$ :

$$\mathcal{W}_i = E_{\Gamma(X)}[\rho_i] = \iiint_{x,y,z} \Gamma(X) \cdot \rho_i \equiv \sum_{x,y,z} \Gamma(X) \cdot \rho_i$$

- 8: Rank cameras in order of highest  $\mathcal{W}_i$  to lowest and display views.  
 9:  $j \leftarrow j + 1$   
 10: Allow user to select the next viewpoint  $\mathcal{V}_j$ .  
 11: Allow user to select the next point  $x_{u,j}$  from  $\mathcal{V}_j$ .  
 12: Compute the SLF for the latest selection  $\Gamma'(X)$ .  
 13:  $\Gamma(X) \leftarrow \Gamma(X) \cdot \Gamma'(X)$ .  
 14: Go to step 7 and repeat until the desired view is selected.
-

The method in greater detail is best understood by reference to Figure 3. The digital archive (12) (shown in Figure 1) contains a plurality of images taken from a scene filled with a number of objects. A display (not shown) linked to the computerized device (not shown) running the computer program (22) displays the representative user interface (23) of the computer program (22), shown in Fig. 4a. The user interface (23) enables the user to view multiple camera angles of a scene in a main viewing area (25), by means of a series of "FUNCTION KEYS" provided by the user interface for example "FORWARD", "PAUSE/PLAY", "BACK", "LIVE", "ZOOM" and so on.

The user interface (23) is operable to permit the user to select an object shown in the main view (25). Thereafter, the system of the present invention: (1) estimates the 3D location of the object selected (30), (2) determines the scene(s) of interest (32) (i.e. the area of the object), (34) estimates the 2D location of the object (36), (4) annotates the 2d location (e.g. red outline) (38), (5) orders the scenes of interest from the various images according to the most probable to relate to the selected object (40), (6) displays the ordered scenes of interest. If optimal localization is achieved (i.e. selected object is visible in desired resolution) (42) the process has been completed, and the optimal image of the object is viewed, associated video is accessed and played back, reversed, viewed in "LIVE" format (44). If the user determines that optimal localization has not been achieved, then the user can access a number of operations that enable further selection of the desired object or area, and thereby the process begins again until on an iterative basis the desired localization is achieved.

For example, user interface (23) displays to the user 15 of the images at any given time, and user wants to find the best 15 images of a particular object, for example a recycling bin. The initial 15 images that are available at random are shown in Figure 7. The recycling bin is selected from the 12<sup>th</sup> image in this set (the top-left image is taken as the first and the bottom right image as the 15th). The algorithm then assigns different rankings to each image, based on the selection of the recycling bin, and reorders the images, displaying the top-ranked as in Figure 8. Taking the expectation of the SLF over the whole space, the location of the desired object is estimated and boxed with a red color. It can be seen that in this iteration, four of the images include the object of interest, and in two, the location has been properly estimated.

This selection process by the user is repeated another two times (to provide the results of Figure 9), and the final ordering of the images is shown in Figure 10. All but one of the images contain the recycling bin. Furthermore, the location of the bin has been determined perfectly in every image. The 13<sup>th</sup> image is an anomaly in this experiment. This image is ranked high, due to a poor calibration at the time of taking the photograph, resulting in incorrectly inferred spatial locations.

It difficult to analyze quantitatively the performance of the system of the present invention. An experiment was performed to quantify the quality of the image-rankings as the number of object selections increases. The same data set as the previous experiment was used. A set of 10 individuals were asked to manually give a score between 0 and 2 to the objects in the images. So the individual would go through every image, and look to see if the objects in question are visible in the image. If the object is completely visible, a score of 2 is given to it; if it is partially visible, a score of 1; and if it is not visible at all, a score of 0 is assigned to the particular object for the particular image. The system is then used to select the particular object and the scores (averaged over all the individuals and objects) are then plotted versus the image-ranks given by the algorithm. It was found that the highest scores given by the individuals participating in this test corresponded to the highest-ranked images as determined in accordance with this invention.

The present invention makes it possible to find specific objects and regions of interest in all of the images by looking at the spatial expectation of the spatial likelihood function. This becomes especially useful in circumstances where a very large number of cameras are available and the man power to sort and search within the images is limited.

The described invention has a multitude of applications. In security and surveillance, it can be used to reduce the number of human monitors, and increase the speed and efficiency of monitoring large environments, like an airport or a casino. For these applications, the system can further be combined with live-streamed video and tracking systems. In advertising, it can be used to create very large-scale databases of images of the item of interest, and make it available to potential clients digitally. As another example, a building company can completely photograph a newly designed building in this fashion.

## REFERENCES

- [1] N. Howe and D. Huttenlocher, "Integrating color, texture, and geometry for image retrieval," in *Proc. IEEE CVPR'2000*, vol. 2, June 13–15, 2000, pp. 239–246.
- [2] S. S. L. Sanghoon Sull; Jeongtaek Oh; Sangwook Oh; Moon-Ho Song, "Relevance  
5 graph-based image retrieval," in *Proc. IEEE ICME'2000*, vol. 2, July 30–August 2, 2000, pp. 713–716.
- [3] A. Kushki, P. Androustos, and A. Venetsanopoulos, "Interactive image retrieval by query fusion," in *Proc. IEEE ICASSP'04*, vol. 3, May 17–21, 2004, pp. 465–468.
- 10 [4] A. Kushki, P. Androustos, K. Plataniotis, and A. Venetsanopoulos, "Retrieval of images from artistic repositories using a decision fusion framework," *IEEE Trans. Image Processing*, vol. 13, no. 3, pp. 277–292, Mar. 2004.
- [5] C. Colombo, A. D. Bimbo, and I. Genovesi, "Interactive image retrieval by color distributions," in *Proc. IEEE Multimedia Computing and Systems*, June 28–July 1, 1998, pp.  
15 255–258.
- [6] R. Aditya and S. Ghosal, "An integrated segmentation technique for interactive image retrieval," in *Proc. IEEE ICIP'2000*, vol. 3, Sept. 10–13, 2000, pp. 762–765.
- [7] W.-H. Lin, R. Jin, and A. Hauptmann, "Web image retrieval re-ranking with relevance model," in *Proc. IEEE WIC'03*, Oct. 13–17, 2003, pp. 242–248.
- 20 [8] P. Muneesawang and L. Guan, "An interactive approach for cbir using a network of radial basis functions," *IEEE Trans. Multimedia*, vol. 6, no. 5, pp. 703–716, Oct. 2004.
- [9] H. Wu, H. Lu, and S. Ma, "Willhunter :interactive image retrieval with multilevel relevance measurement," in *Proc. IEEE ICPR'04*, vol. 2, Aug. 23–26, 2004, pp. 1009–1012.
- [10] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed.  
25 Cambridge, UK: Cambridge University Press, 2004.

- [11] R. I. Hartley, E. Hayman, L. de Agapito, and I. Reid, "Camera calibration and the search for infinity," *Computer Vision. The Proceedings of the Seventh IEEE International Conference on*, vol. 1, pp. 510–517, Sept. 1999.
- [12] O. Faugeras, "From geometry to variational calculus: theory and applications of three-dimensional vision," *Computer Vision for Virtual Reality Based Human Communications, Proceedings., IEEE and ATR Workshop on*, pp. 52–71, Jan. 1998.
- [13] R. I. Hartley, "Self-calibration of stationary cameras," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 5–23, 1997.
- [14] E. Bayro-Corrochano and B. Rosenhahn, "A geometric approach for the analysis and computation of the intrinsic camera parameters," *Pattern Recognition*, vol. 35, pp. 338–347, Jan. 2002.
- [15] F. Pedersini, A. Sarti, and S. Tubaro, "Multi-camera systems," *IEEE Signal Processing Magazine*, vol. 16, no. 3, pp. 55–65, May 1999.
- [16] F. Pedersini and S. Tubaro, "Multi-camera parameter tracking," *Vision, Image and Signal Processing, IEE Proceedings*, vol. 148, no. 1, pp. 70–77, Feb. 2001.
- [17] Z. Zhang and V. Schenk, "Self-maintaining camera calibration over time," *Computer Vision and Pattern Recognition. Proceedings., IEEE Computer Society Conference on*, pp. 231–236, June 1997.
- [18] C. T. Huang and O. R. Mitchell, "Dynamic camera calibration," *Computer Vision, Proceedings., International Symposium on*, vol. 3, pp. 2165–2170, May 2001.
- [19] T. Pajdla and V. Hlaváček, "Camera calibration and euclidean reconstruction from known observer translations," *Computer Vision and Pattern Recognition, Proceedings., IEEE Computer Society Conference on*, pp. 421–426, June 1998.
- [20] M. Devy, V. Garic, and J. Orteu, "Camera calibration from multiple views of a 2d object, using a global nonlinear minimization method," *Intelligent Robots and Systems, 1997. IROS '97., Proceedings of the 1997 IEEE/RSJ International Conference on*, vol. 3, pp. 1583–1589, Sept. 1997.

[21] C. Wiles and A. Davison, "Calibrating a multi-camera system for 3d modelling," *Multi-View Modeling and Analysis of Visual Scenes, 1999. (MVIEW '99) Proceedings. IEEE Workshop on*, pp. 29 – 36, June 1999.

[22] F. Pedersini, A. Sarti, and S. Tubaro, "Accurate and simple geometric calibration of  
5 multi-camera systems," *Signal Processing*, vol. 77, pp. 309–334, 1999.

[23] P. Aarabi, "Localization-based sensor validation using the kullbackleibler divergence,"  
*IEEE Trans. Syst., Man, Cybern.*, vol. 32, no. 2, pp. 1007–1016, Apr. 2004.

**CLAIMS**

1. A method of searching scenes in images, retrieving images, and/or localizing objects in images, characterized by:
  - (a) obtaining a plurality of images from one or more cameras, the plurality of images including at least two images including a view of a single area or object shown in a scene;
  - (b) selecting a particular area of interest or object of interest in the scene; and
  - (c) iteratively establishing a sub-set of the plurality of images that are probably of interest for viewing the particular area of interest or object of interest by:
    - (i) determining a probability distribution of the plurality of images based on location data and data regarding the geometry of an environment of the scene established for the area of interest or object of interest;
    - (ii) refining the probability distribution by obtaining user input regarding:
      - (A) one or more of the sub-set of plurality images that the user considers to be most relevant from the current sub-set of the plurality of images; and
      - (B) selection of the particular area of interest or object of interest in the one or more most relevant images; and
    - (iii) updating the sub-set of the plurality of images based on the user input of (ii).
2. The method of claim 1, characterized by the further step of calibrating the one or more cameras to enable a transformation whereby three-dimensional coordinates of points within the field of view of the one or more cameras are mapped to a two-dimensional plane.



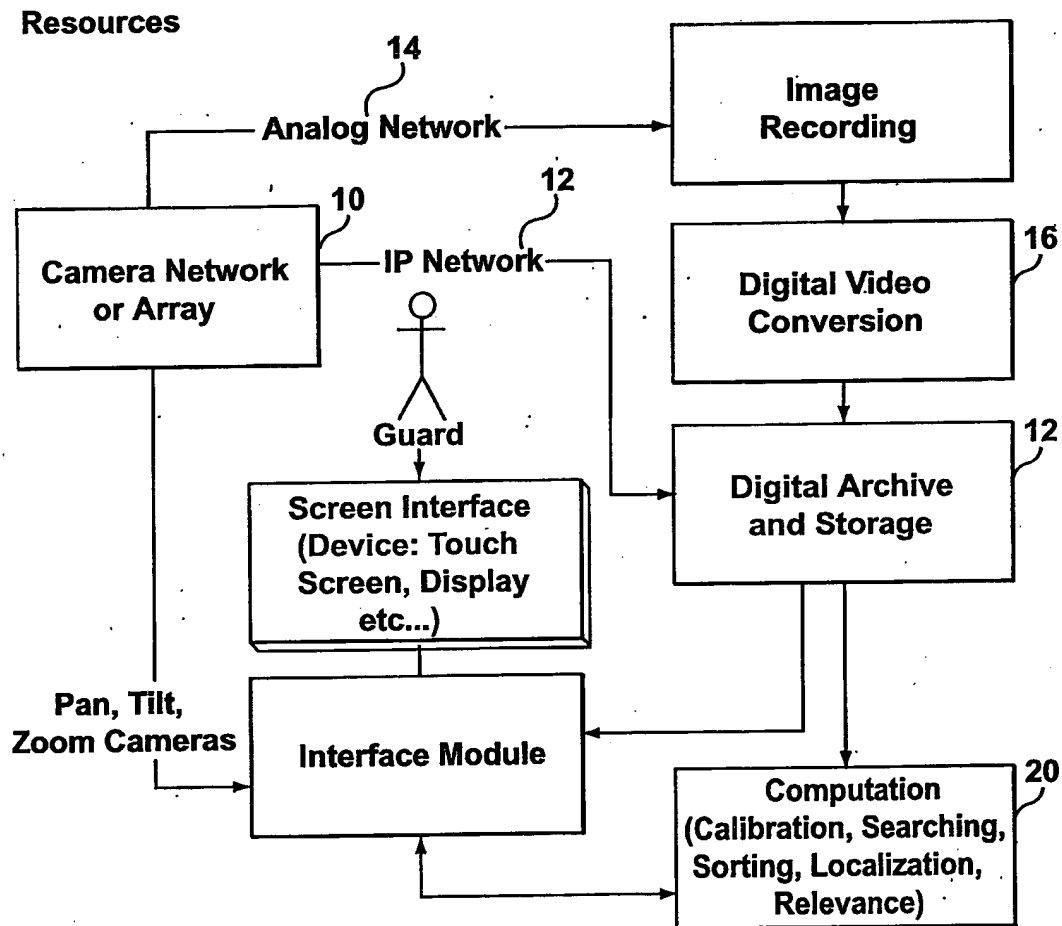
3. The method of claim 1, characterized by the further step of generating a spatial reliability function for each of the one or more cameras, and assigning the corresponding spatial reliability function to each of the one or more cameras.
4. The method of claim 3, characterized by the further step of generating a spatial likelihood function for the selected particular area of interest or object of interest, and combining each of the applicable spatial reliability functions with the spatial likelihood function, so as to establish the probability distribution.
5. The method of claim 4, characterized by the combination of the applicable spatial reliability function with the spatial likelihood function consisting of point-by-point multiplication and normalization.
6. The method of claim 3, characterized in that the spatial reliability function is adjusted for decrease of reliability over distance.
7. The method of claim 1, characterized by the further step of ranking the sub-set of the plurality of images according to relevance and displaying such ranking to the user.
8. A method of searching scenes in images, retrieving images, and/or localizing objects in images, characterized by:
  - (a) displaying one or more images obtained from one or more cameras calibrated to provide location data and data regarding the geometry of an environment of a scene;
  - (b) a user selecting an object or area of interest in the scene;
  - (c) accessing a plurality of images associated with the object or area;
  - (d) estimating the three-dimensional location of the object or area, and estimating the two-dimensional location of the object or area, so as to define location data for the object or area;

- (e) determining whether the object or area is present in the plurality of images, so as to establish a first sub-set of the plurality of images consisting of images that include the object or area;
  - (f) determining one or more second sub-sets of the plurality of images, optionally including images from the first sub-set of images, iteratively established to be probably of interest for viewing the object or area by:
    - (i) determining a probability distribution of the plurality of images based on the location data and data regarding the geometry of the environment of the scene established for the area of interest or object of interest;
    - (ii) refining the probability distribution by obtaining user input regarding:
      - (A) one or more of the images of the second sub-set of images that the user considers to be most relevant from the current second sub-set of the plurality of images; and
      - (B) selection of the particular area of interest or object of interest in the one or more most relevant images; and
    - (iii) updating the current second sub-set of the plurality of images based on the user input of (ii).
9. A system for searching scenes in images, retrieving images, and localizing objects in images, characterized in that the system includes:
- (a) at least one camera calibrated to provide location data and data regarding the geometry of an environment of a scene, for a plurality of images, the plurality of images including at least two images including a view of a single area or object shown in a scene; and
  - (b) a computer linked to the at least one camera, the computer including a computation utility, the computation utility being operable on the computer to:

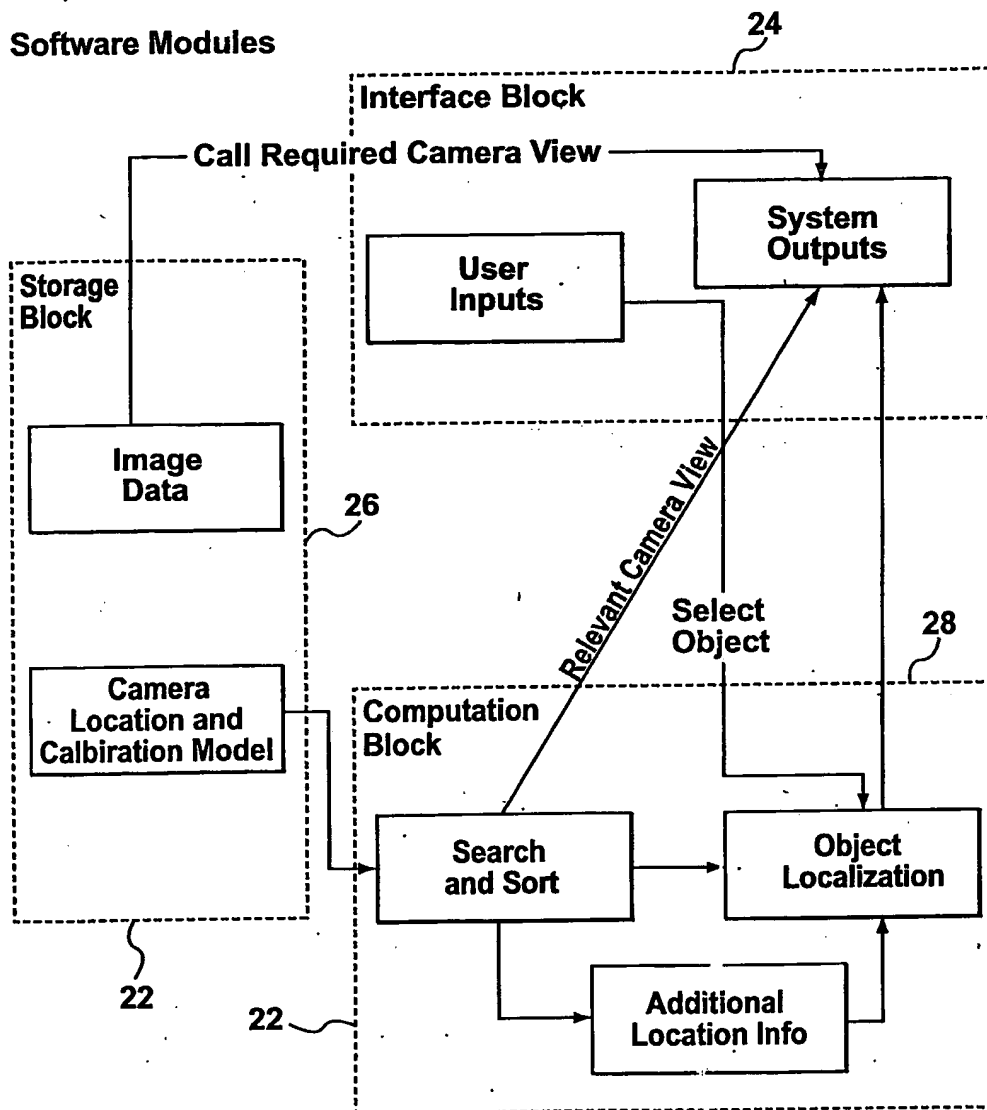
- (i) select a particular area of interest or object of interest in the scene; and
  - (ii) iteratively establish a sub-set of the plurality of images that are probably of interest for viewing the particular area of interest or object of interest by:
    - (A) determining a probability distribution of the plurality of images based on location data and data regarding the geometry of the environment of the scene established for the area of interest or object of interest;
    - (B) refining the probability distribution by obtaining user input regarding:
      - (I) one or more of the sub-set of plurality images that the user considers to be most relevant from the current sub-set of the plurality of images; and
      - (II) selection of the particular area of interest or object of interest in the one or more most relevant images; and
    - (C) updating the sub-set of the plurality of images based on the user input of (B).
10. A computer program for searching scenes in images, retrieving images, and localizing objects in images, characterized in that the computer program includes instructions operable on a computer to:
- (a) enable an interface with at least one camera calibrated to provide location data and data regarding the geometry of an environment of a scene, the plurality of images including at least two images including a view of a single area or object shown in a scene; and
  - (b) provide a computation utility operable to enable a user to:
    - (i) select a particular area of interest or object of interest in the scene; and

- (ii) iteratively establish a sub-set of the plurality of images that are probably of interest for viewing the particular area of interest or object of interest by:
  - (A) determining a probability distribution of the plurality of images based on location data and data regarding the geometry of an environment of the scene established for the area of interest or object of interest;
  - (B) refining the probability distribution by obtaining user input regarding:
    - (I) one or more of the sub-set of plurality images that the user considers to be most relevant from the current sub-set of the plurality of images;
    - (II) selection of the particular area of interest or object of interest in the one or more most relevant images; and
  - (C) updating the sub-set of the plurality of images based on the user input of (B);

1 / 12

**FIG. 1**

2 / 12

**FIG. 2**

3 / 12

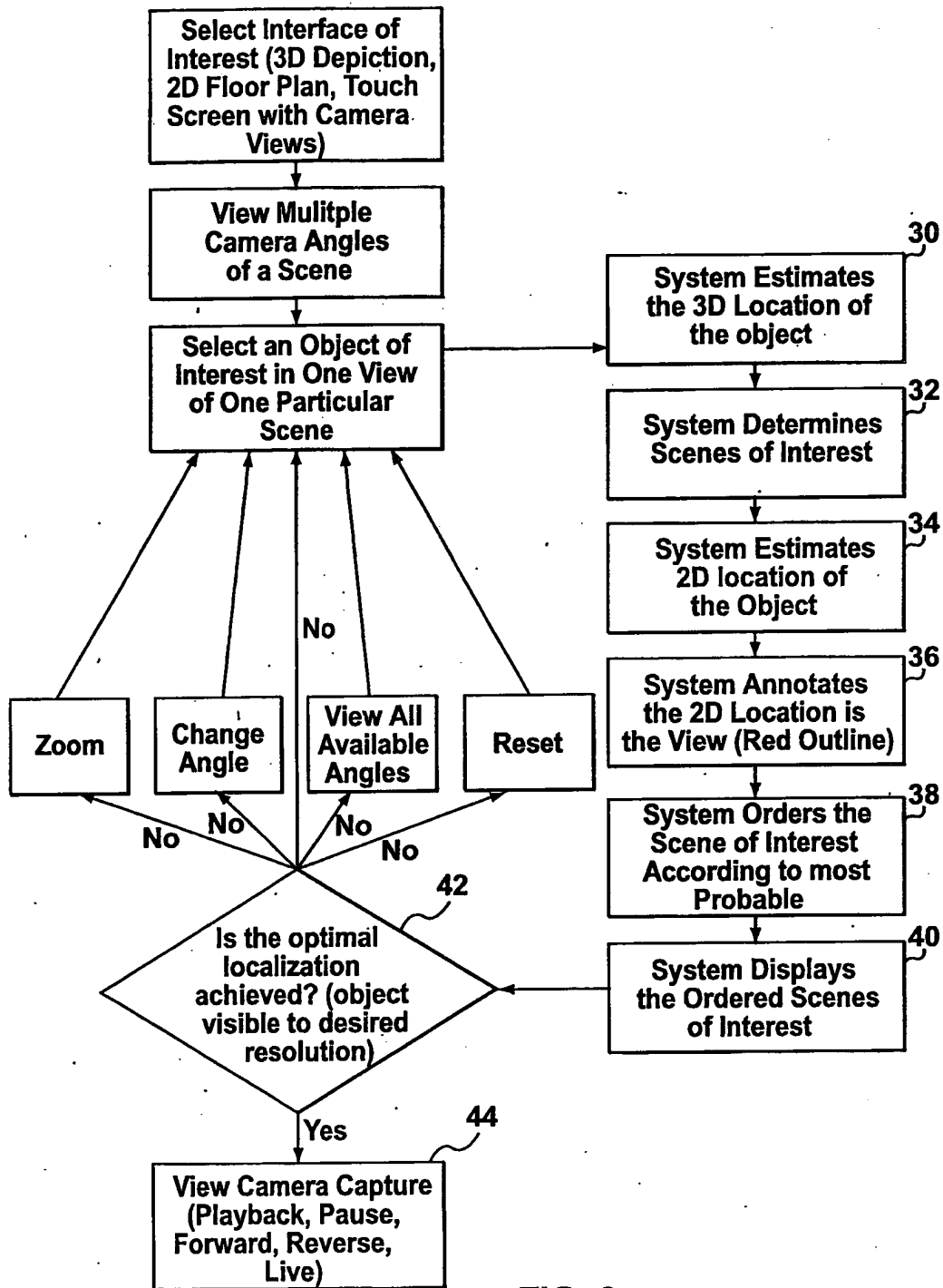
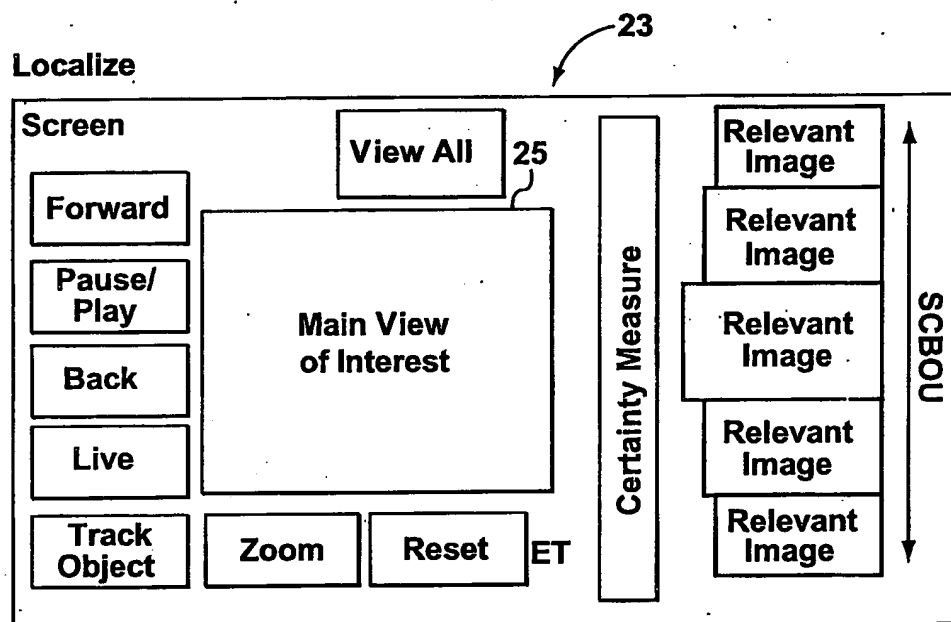


FIG. 3

4 / 12

**Preferred User Interface****FIG. 4a**



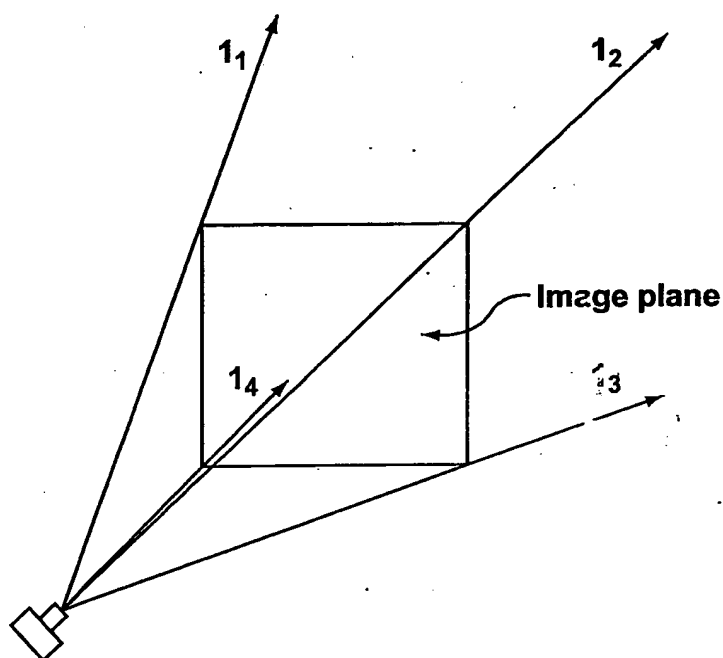
5 / 12

Display All

23

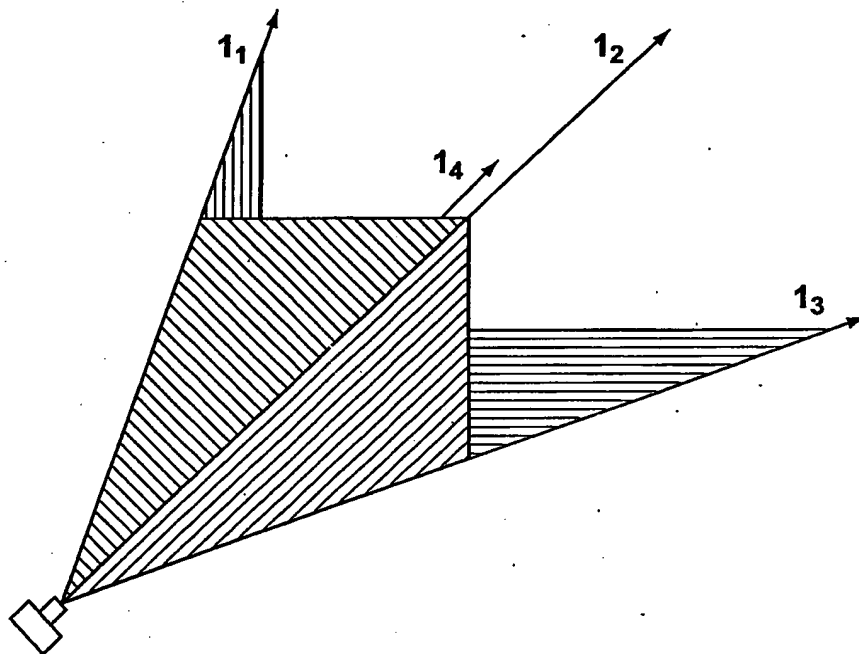
Screen							
View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene
View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene
View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene
View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene
View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene
View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene
View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene
View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene	View of Scene
Next				Previous			

**FIG. 4b**

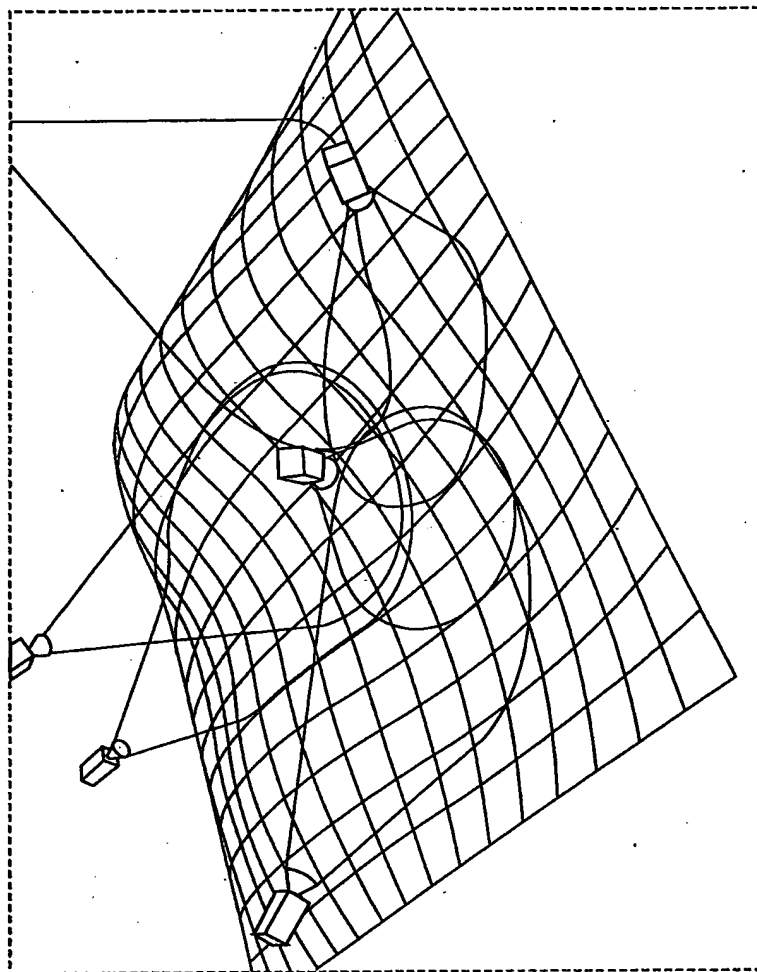


**FIG. 5a**

7 / 12

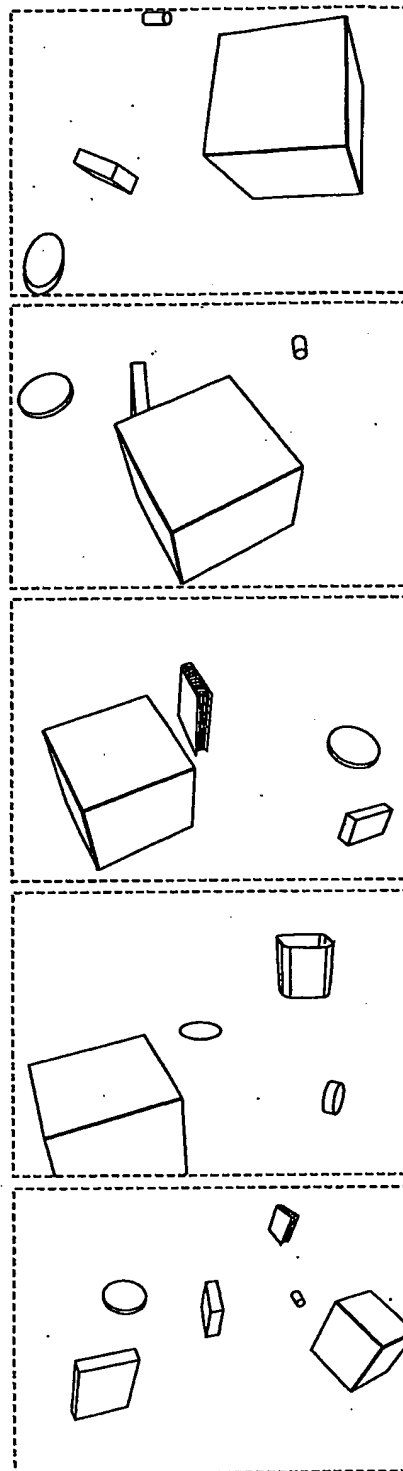


**FIG. 5b**



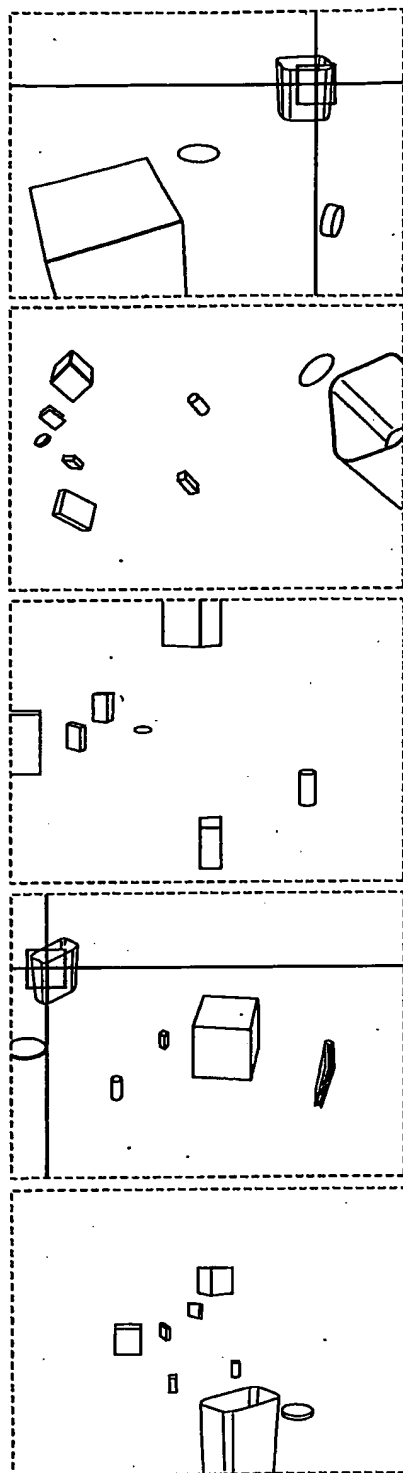
**FIG. 6**

9 / 12



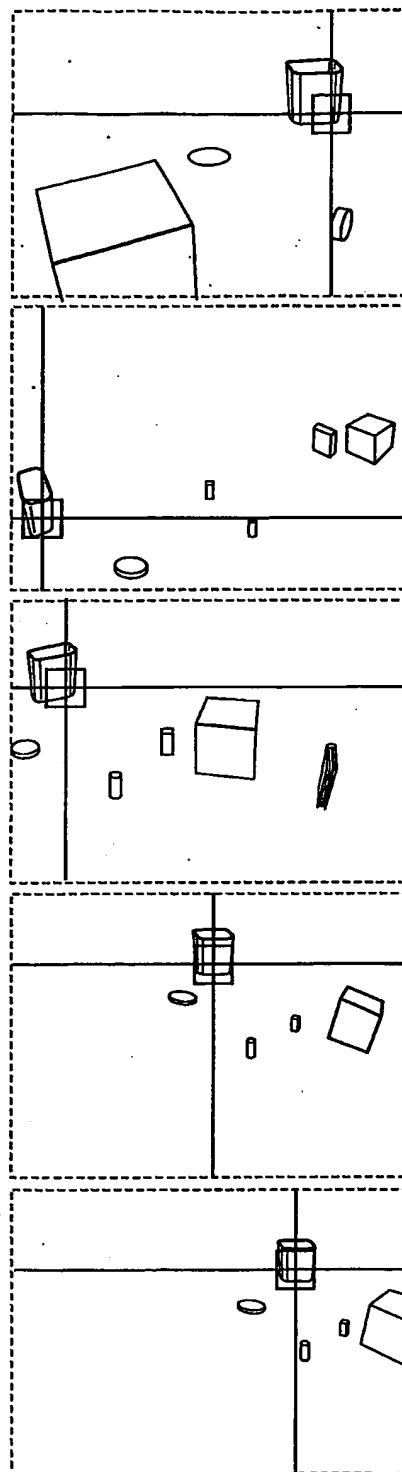
**FIG. 7**

10 / 12



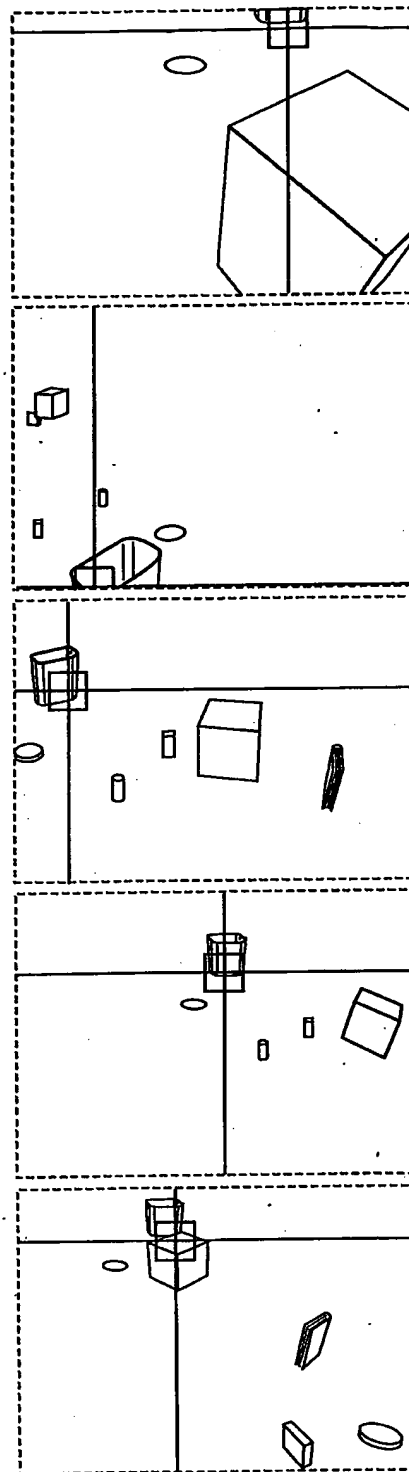
**FIG. 8**

11 / 12



**FIG. 9**

12 / 12



**FIG. 10**



# INTERNATIONAL SEARCH REPORT

Corrected version on date of mailing  
13 December 2005 (13-12-2005)

International application No.  
PCT/CA2005/001093

<b>A CLASSIFICATION OF SUBJECT MATTER</b> IPC(7): G06T 7/00, G06T 7/60, G06F 17/30  According to International Patent Classification (IPC) or to both national classification and IPC				
<b>B. FIELDS SEARCHED</b>  Minimum documentation searched (classification system followed by classification symbols)  IPC(7): G06T 7/00, G06T 7/60 and G06F 17/30  Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched				
Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used) Canadian database, IEEE and Delphion, Keywords: search*, scene, image, retrieval, localizing, camera, area, sub-set, calibrating, three-dimensional, coordinates, two-dimensional, probability, distribution, geometry, environment, user, probability, spatial, reliability and				
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>				
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
X	US 5,696,964 (NEC RESEARCH INSTITUTE, INC) 09 December 1997 (09-12-1997)	1,8 to 10		
Y	columns 4 and 6 column 7, lines 16 to 38 columns 11 and 12	2 to 7		
Y	Aarabi, 'Self-localizing dynamic microphone arrays', 2002, IEEE Transaction on Systems, Man, and Cybernetics Part C (32) 4 pp. 474-484. see whole document	3 to 6		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of <input checked="" type="checkbox"/> See patent family annex.				
<table border="0"> <tr> <td>           * Special categories of cited documents            "A" document defining the general state of the art which is not considered to be of particular relevance            "B" earlier application or patent but published on or after the international filing date            "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)            "O" document referring to an oral disclosure, use, exhibition or other means            "P" document published prior to the international filing date but later than the priority date claimed         </td> <td>           "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention            "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone            "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art            "Z" document member of the same patent family         </td> </tr> </table>			* Special categories of cited documents "A" document defining the general state of the art which is not considered to be of particular relevance "B" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "Z" document member of the same patent family
* Special categories of cited documents "A" document defining the general state of the art which is not considered to be of particular relevance "B" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "Z" document member of the same patent family			
Date of the actual completion of the international search 04 November 2005 (04-11-2005)		Date of mailing of the international search report 14 November 2005 (14-11-2005)		
Name and mailing address of the ISA/CA Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 001(819)953-2476		Authorized officer  Gilbert Jean (819) 934-5148		

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/CA2005/001093

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6,574,616 (INDEX STOCK IMAGERY, INC) 03 June 2003 (03-06-1997) column 5, lines 44 to 67 column 6	1 and 7 to 10
Y	column 7, lines 1 to 25 column 8	2 to 6

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.  
PCT/CA2005/001093

Patent Document Cited in Search Report	Publication Date	Patent Family Member(s)	Publication Date
US5696964	09-12-1997	CN1094219C C	13-11-2002
		DE69723346D D1	14-08-2003
		DE69723346T T2	09-06-2004
		EP0802489 A2	22-10-1997
		JP3053003B2 B2	19-06-2000
US6574616	03-06-2003	NONE	